# SkELL: A discovery-based Chinese learning platform

Simon Smith (Coventry University)

Miloš Jakubíček (Lexical Computing)

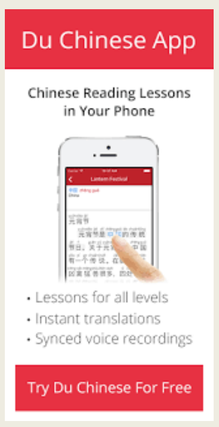
## 一 Outline

- Chinese learning
- ICT for Chinese
- What's missing (corpora)
- Planned implementation
- Optimize gramrels (rules) for Chinese
- Implement multi-level segmentation

## 二 Importance of Chinese

- Mandarin Chinese has most native speakers of all languages.
- GCSE entries up 18% in 2015, despite overall decline language exam entries (Guardian 2015).
- Popularity at primary school (Chen 2014)
- Mandarin mastery important for careers (Lo 2016)
- Trade with China
- Confucius Institutes

## 三 Existing ICT provision






## 四

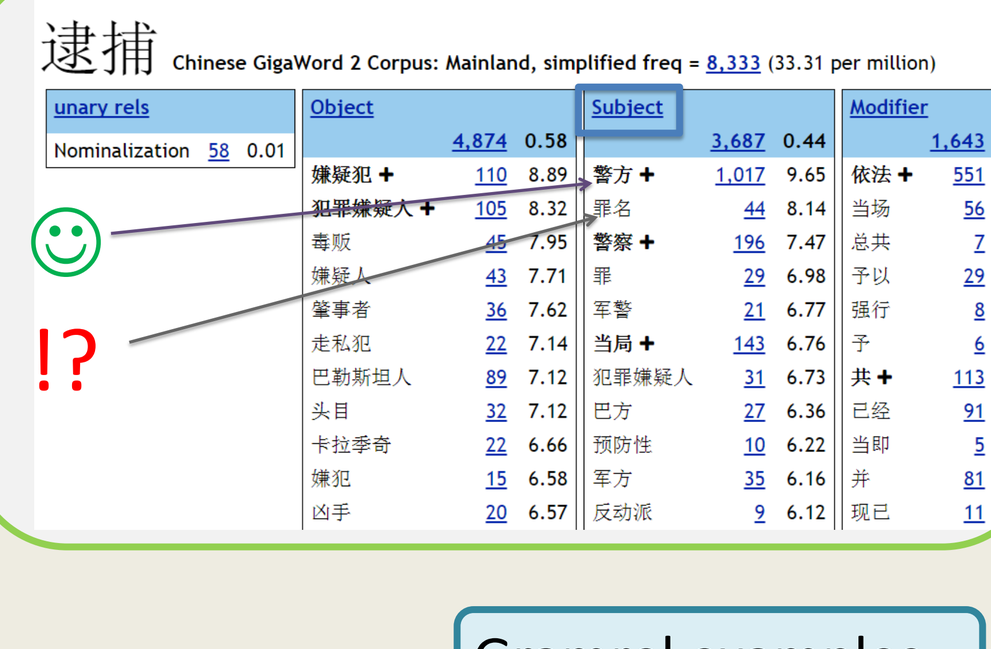
Sketch Engine

- Acclaimed corpus analysis platform
- English learner dictionaries
- Resources for many languages
- Suitable for expert users
- **Can** be used by teachers/learners

SkELL

- Sketch Engine for Language Learning
- Free on web
- English, Czech & Russian currently available
- German & Italian soon
- Suitable for teachers/learners
- Extensible to other languages

## 五 First, an English example: studying collocation with SkELL Word sketch




## 六 Next, a Chinese example: studying collocation of 逮捕 *arrest*




## 七 Gramrel examples

**Simple VO rule for English**

```
1: "V" "(DET|NUM|ADJ|ADV|N)"* 2:"N"
```

**Object-fronting direct object rule for Chinese**

```
[word="把"|word="将"] NP adv{0,2}
1:"VV" (particle|prep)? NP1 noun
```

## 八 Chinese segmentation privileges longer words

- Longest match procedure (Tseng & Chen 2002)
- OOV words often long (Ma & Chen 2005)
- Sun et al (1998) use MI segmentation without lexicon
- MI threshold is parametrizable
- Different word lengths could be favoured

[ 中 国 人 ]
[ 中 国 ][ 人 ]
[ 中 ][ 国 ][ 人 ]

Lexicography standard

Learners: word level

Learners: intra-word

## References

Chen, T. (2014). Teaching Chinese as a Foreign Language at Primary School in England. Quarterly Journal of Chinese Studies, 2(4), 67-83.

Guardian (2015). *GCSE results: fall in numbers taking foreign languages 'a cause for concern'*. Available at https://www.theguardian.com/education/2015/aug/20/gcse-results-fall-numbers-foreign-languages [2 January 2017]

Lo, L. (2016). Challenges faced by Cantonese speakers in a UK university Mandarin course. *Innovative language teaching and learning at university: enhancing participation and collaboration*, 139-145.

Ma, W. Y., & Chen, K. J. (2005). Design of CKIP Chinese word segmentation system. *Chinese and Oriental Languages Information Processing Society*, *14*(3), 235-249.

Sun, M., Shen, D., & Tsou, B. K. (1998). Chinese word segmentation without using lexicon and hand-crafted training data. In *Proceedings of the 17th international conference on Computational linguistics-Volume 2* (pp. 1265-1271). Association for Computational Linguistics.

Tseng, H., & Chen, K. J. (2002). Design of Chinese morphological analyzer. In: *Proceedings of the first SIGHAN workshop on Chinese language processing* (pp. 1-7). Association for Computational Linguistics.